Bridging Model and Crop Legumes through Comparative Genomics

Hongyan Zhu*, Hong-Kyu Choi, Douglas R. Cook, and Randy C. Shoemaker

Department of Plant and Soil Sciences, University of Kentucky, Lexington, Kentucky 40546 (H.Z.); Department of Plant Pathology, University of California, Davis, California 95616 (H.-K.C., D.R.C.); and Corn Insect and Crop Genetics Research Unit, United States Department of Agriculture, Agricultural Research Service, Iowa State University, Ames, Iowa 50011 (R.C.S.)

The Fabaceae, or legumes, constitute the third largest family of flowering plants, comprising more than 650 genera and 18,000 species (Polhill and Raven, 1981). Economically, legumes represent the second most important family of crop plants after Poaceae (grass family), accounting for approximately 27% of the world's crop production (Graham and Vance, 2003). On a worldwide basis, legumes contribute about one-third of humankind's protein intake, while also serving as an important source of fodder and forage for animals and of edible and industrial oils. One of the most important attributes of legumes is their unique capacity for symbiotic nitrogen fixation, underlying their importance as a source of nitrogen in both natural and agricultural ecosystems. Legumes also accumulate natural products (secondary metabolites) such as isoflavonoids that are beneficial to human health through anticancer and other health-promoting activities (Dixon and Sumner, 2003).

The legumes are highly diverse and can be divided into three subfamilies: Mimosoideae, Caesalpinioideae, and Papilionoideae (Doyle and Luckow, 2003). Of these, the Papilionoideae subfamily contains nearly all economically important crop legumes, including soybean (*Glycine max*), peanut (*Arachis hypogaea*), mungbean (*Vigna radiata*), chickpea (*Cicer arietinum*), lentil (*Lens culinaris*), common bean (*Phaseolus vulgaris*), pea (*Pisum sativum*), and alfalfa (*Medicago sativa*). With the notable exception of peanut, all these important crop legumes fall into two Papilionoid clades, namely, Galegoid and Phaseoloid, which are often referred to as cool season and tropical season legumes, respectively (Fig. 1). Despite their close phylogenetic relationships, crop legumes differ greatly in their genome size, base chromosome number, ploidy level, and self-compatibility (Table I). Nevertheless, earlier studies indicated that members of the Papilionoideae subfamily exhibited extensive genome conservation based on comparative genetic mapping (Weeden et al., 1992; Menancio-Hautea et al., 1993). To establish a unified genetic system for legumes, two legume species in the Galegoid clade, *Medicago truncatula* and *Lotus japonicus*, which belong to the tribes Trifolieae and Loteae, respectively, were selected as model systems for studying legume genomics and biology (Cook, 1999; Stougaard, 2001). Unlike many of the major crop legumes, *M. truncatula* and *L. japonicus* are of small genome size, amenable to forward and reverse genetic analyses, and well suited for studying biological issues important to the related crop legume species.

An immediate goal of legume genomics is to transfer knowledge between model and crop legumes. Accordingly, an in-depth understanding of conservation of genome structure among legume species is a prerequisite to achieving this goal. The idea that conserved genome structure can facilitate transfer of knowledge among related plant species is best addressed in grasses in which genome macrosynteny and microsynteny have been extensively maintained (Bennetzen, 2000; Devos and Gale, 2000). These studies, however, also revealed many exceptions to the conserved synteny, with frequent local genic rearrangements including gene inversion, duplication, translocation, and insertion/deletion. Although the degeneracy of local genome microstructure has been widely documented, it is less clear the extent to which such alterations to genome microstructure contribute to the divergence of genome function. In this review, we focus on recent results of comparative genome analysis between model and crop legumes, and also highlight the recent successes of using comparative genomics tools for cross-species gene isolation.

DUPLICATIONS THAT SHAPE THE LEGUME GENOMES

It has been estimated that upwards of 80% of all angiosperms are likely to have a polyploid origin (Masterson, 1994). It is unlikely that legumes are an exception to this. Soybean, for example, has long been known to be an ancient polyploid with putative homoeologous chromosomal regions readily identified by genetic mapping (Shoemaker et al., 1996; Lee et al., 1999, 2001) and by characterization of homoeologous bacterial artificial chromosome (BAC) clones (Foster-Hartnett et al., 2002; Yan et al., 2003). And recently, segmental duplications within the soybean genome were visualized by fluorescence in situ

* Corresponding author; e-mail hzhu4@uky.edu; fax 859–323–1077.

www.plantphysiol.org/cgi/doi/10.1104/pp.104.058891.

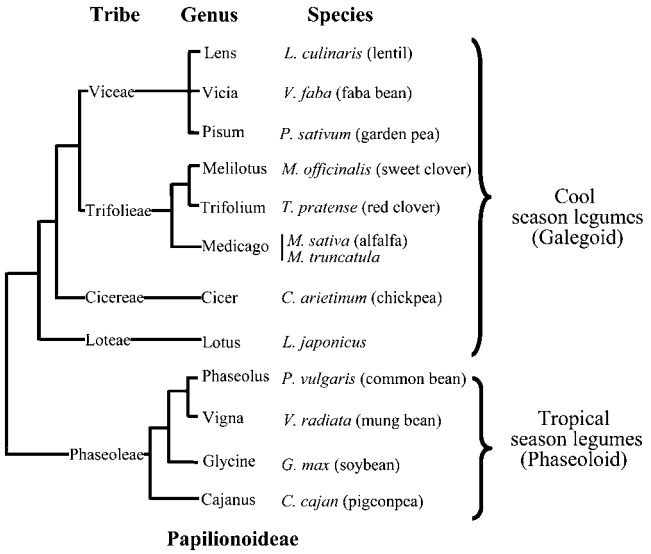

Figure 1. Dendrogram depicting phylogenetic relationships of Papilionoideae legumes. (Figure reprinted from Choi et al. [2004b], based on figure 5 of Doyle and Luckow [2003].)

hybridization of BACs (Pagel et al., 2004). Segmental duplications also were identified in the *M. truncatula* and *L. japonicus* genomes through high-throughput genome sequencing (Zhu et al., 2003; N. Young, personal communication).

Extensive expressed sequence tag (EST) collections exist for two legumes representing distinct phylogenetic clades, soybean (tropical legumes) and *M. truncatula* (cool season legumes). Much of a genome's evolutionary history can be read in these transcripts. Schlueter et al. (2004) identified duplicate transcripts from the EST collections of both of these legumes and estimated genetic distances of the pairs using synonymous substitution measurements. It was estimated that soybean probably underwent two major genome duplications events: one at 15 million years ago (MYA) and another at 44 MYA. A genome duplication event also was estimated to have occurred in *M. truncatula* at approximately 58 MYA. A subsequent analysis using a multigene approach concluded that the more ancient duplication events probably represent a single event that occurred before soybean and Medicago diverged (Pfeil et al., 2005). If this is true, then approximately 7,000 other legumes share the same genome duplication event (Pfeil et al., 2005).

Genome duplications often are followed by gene loss, rearrangements, tandem gene or segmental duplications, and divergence of duplicated gene sequences. All of these events are involved in the process of diploidization (Ohno, 1970) and complicate the interpretation of comparative genomic data.

MACROSYNTENY AMONG PAPILIONOID LEGUMES

Macrosynteny generally refers to conserved gene order between species revealed by comparative genetic mapping of common DNA markers or in silico mapping of homologous sequences. Early comparative studies of legume genomes were focused on closely related species of the same genus or tribe, based primarily on comparative mapping of common RFLP markers. Weeden et al. (1992) first reported conserved gene order between pea and lentil, accounting for approximately 40% of the lentil genome. Later, Menancio-Hautea et al. (1993) demonstrated that mungbean and cowpea (*Vigna unguiculata*) also exhibited a high degree of linkage conservation, whereas chromosomal rearrangements have occurred since the divergence of the two species. Comparative mapping among mungbean, common bean, and soybean in the Phaseoleae tribe indicated that mungbean and common bean linkage groups were highly conserved, but synteny with soybean was limited only to the short linkage blocks (Boutin et al., 1995). A more recent study, however, using Arabidopsis (*Arabidopsis thaliana*) as

Table I. *Chromosome number and genome size of major model and crop legumes*[a]

Tribe	Genus	Species	Chromosome No.	Genome Size	Self-Compatibility
				Mb/1C	
Trifolieae	Medicago	*M. truncatula* (barrel medic)	2*n* = 2x = 16	466	Selfing
		Alfalfa	2*n* = 4x = 32	1,715	Outcrossing
	Trifolium	*Trifolium pretense* (red clover)	2*n* = 2x = 14	637	Outcrossing
		Trifolium repens (white clover)	2*n* = 4x = 32	956	Outcrossing
	Melilotus	*Melilotus officinalis* (sweet clover)	2*n* = 2x = 16	1,103	Outcrossing
Viceae	Pisum	Garden pea	2*n* = 2x = 14	4,337	Selfing
	Vicia	*Vicia faba* (faba bean)	2*n* = 2x = 12	13,059	Selfing
	Lens	Lentil	2*n* = 2x = 14	4,116	Selfing
Cicereae	Cicer	Chickpea	2*n* = 2x = 16	931	Selfing
Loteae	Lotus	*L. japonicus*	2*n* = 2x = 16	466	Selfing
Phaseoleae	Phaseolus	Common bean	2*n* = 2x = 22	588	Selfing
	Vigna	Mungbean	2*n* = 2x = 22	515	Selfing
	Glycine	Soybean	2*n* = 4x = 40	1103	Selfing
	Cajanus	*Cajanus cajan* (pigeon pea)	2*n* = 2x = 22	858	Selfing

[a]Data were from the Plant DNA C-values Database (http://www.rbgkew.org.uk/cval/homepage.html).

a bridging species revealed that homoeologous segments of soybean chromosomes showed a higher degree of synteny with chromosomes of common bean and mungbean than previously thought (Lee et al., 2001).

The most in-depth analysis of legume macrosynteny recently was reported by Choi et al. (2004a, 2004b) using *M. truncatula* as a central point of comparison. This research took advantage of abundant EST sequence information from the model legume *M. truncatula* to develop cross-species genetic markers where locus orthology was tested through phylogenetic analysis. Gene-specific PCR primers were designed to anneal to highly conserved exon sequences that span predicted introns, which allowed for efficient PCR amplification across species and for developing single nucleotide polymorphism markers for linkage mapping in multiple taxa. These putatively orthologous markers were mapped in *M. truncatula*, alfalfa, pea, mungbean (Choi et al., 2004a, 2004b), and chickpea (H. Zhu and D. Cook, unpublished data). In addition, 60 markers developed based on homology to mapped genetic markers of soybean were mapped in *M. truncatula.* Furthermore, the macrosyntenic relationship between *M. truncatula* and *L. japonicus* was evaluated based on 63 pairs of sequenced BAC clones, representing putatively orthologous loci with known genetic position in both species.

A simplified consensus comparative map of eight legume species is shown in Figure 2. As expected, the degree of synteny is correlated with the phylogenetic distance of these legume species. *M. truncatula* and alfalfa share highly conserved nucleotide sequences and exhibit nearly perfect synteny between the two genomes (Choi et al., 2004a). Although the pea genome is approximately 10 times larger than that of *M. truncatula* and has one less chromosome, the colinearity of genes is also remarkably conserved between the two genomes, with major evident differences being inferred interchromosomal rearrangements (Choi et al., 2004b). It was suggested that chromosomal rearrangements involving Medicago (alfalfa and *M. truncatula*) chromosome 6 might be responsible for the difference in chromosome number between Medicago and pea (Choi et al., 2004b; Kalo et al., 2004). Interestingly, the same chromosome seems to have also been associated with the interchromosomal rearrangements between *M. truncatula* and chickpea (H. Zhu and D. Cook, unpublished data). Even though *M. truncatula* and chickpea share the same base chromosome number of 8, one-to-one relationships do not hold true for *M. truncatula* linkage groups 5 and 6 and

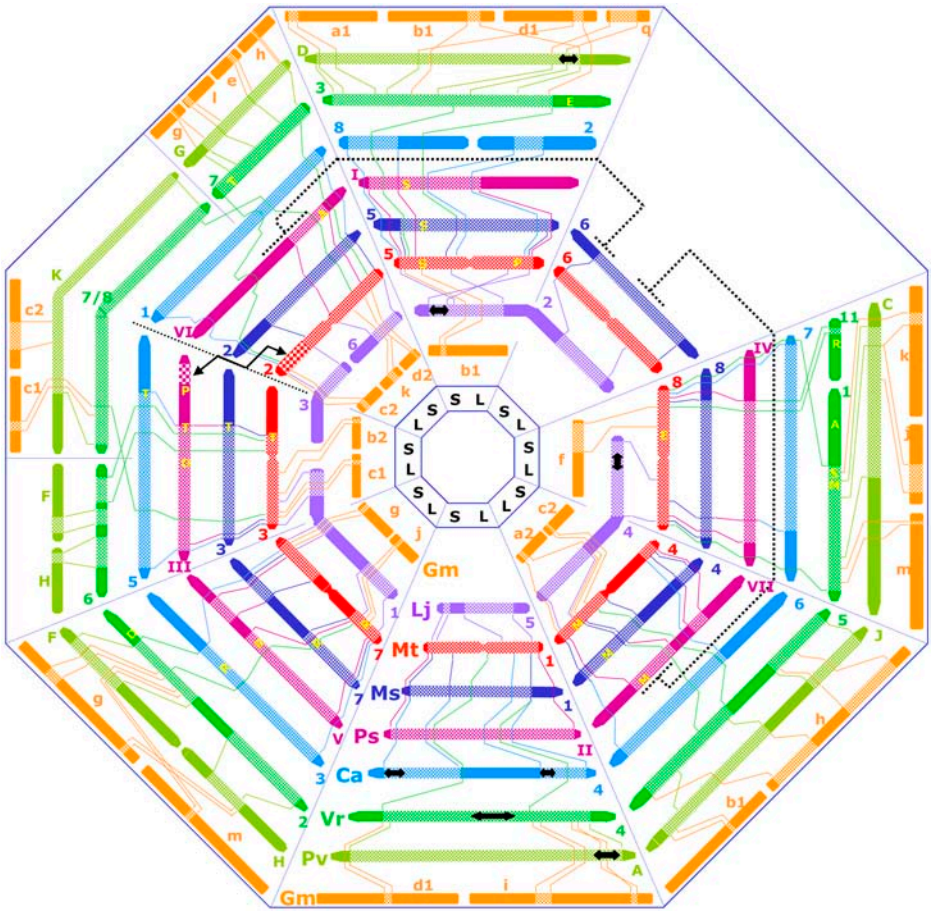

Figure 2. A simplified consensus map for eight legume species. The figure is based on figure 5 of Choi et al. (2004b) with modification. Mt, *M. truncatula*; Ms, alfalfa; Lj, *L. japonicus*; Ps, pea; Ca, chickpea; Vr, mungbean; Pv, common bean; Gm, soybean. S and L denote the short and long arms of each chromosome in *M. truncatula*. Syntenic blocks are drawn to scale based on genetic distance.

chickpea linkage groups 2 and 8. In particular, *M. truncatula* LG5 can be aligned with chickpea LG2 and LG8. Similarly, the genomes of *M. truncatula* and *L. japonicus* also are highly syntenic, but the synteny often is punctuated by chromosomal rearrangements, reflecting the difference of chromosome numbers between the two genomes.

By contrast, macrosyntenic relationships between *M. truncatula* and Phaseoloid legumes were more complicated and less informative. Twenty-nine of the 38 (approximately 76%) markers mapped between *M. truncatula* and mungbean revealed evidence of conserved gene order, whereas the remaining markers mapped to nonsyntenic positions. Similarly, 23 of the 60 mapped markers identified 11 syntenic blocks between *M. truncatula* and soybean. The finding that synteny was limited only to small genetic intervals between more distantly related legumes suggests correlation between the frequency of chromosomal rearrangement and divergence time, which also is reflected by the differences in chromosome number between Galegoid and Phaseoloid legumes. In the case of soybean, duplication (polyploidization) followed by gene loss and segmental reshuffling (diploidization) may make it difficult to identify lengthy stretches of syntenic chromosome segments between soybean and related legumes.

MICROSYNTENY AMONG *M. TRUNCATULA*, *L. JAPONICUS*, AND SOYBEAN

In contrast with macrosynteny, microsynteny often refers to conserved gene content and order at sequence level over a short, physically defined DNA contig. Nearly all the interspecies analyses of microsynteny reported so far have been based on comparisons of a limited number of specific regions, and the conclusions drawn therein may not be extended to the global level because genome microstructure is highly dynamic and the level of conservation varies with different parts of a genome.

Yan et al. (2003) estimated the level of microsynteny between *M. truncatula* and soybean using a hybridization strategy involving BAC contigs. Twenty-seven of 50 soybean contigs (54%) were shown to possess some level of microsynteny with *M. truncatula*. Sequence analysis of regions around the putatively orthologous apyrase genes between *M. truncatula* and soybean also revealed conserved gene order, with at least 6 genes in common over 70 kb (Cannon et al., 2003). Similar comparison was conducted between the *rgh1* locus of soybean and the putatively orthologous region of *M. truncatula* (Choi et al., 2004b). From a total of 29 distinct genes identified in *M. truncatula* and soybean within the syntenic interval, 14 (approximately 48%) were conserved between the two genomes.

More extensive analysis of microsynteny between *M. truncatula* and *L. japonicus* was facilitated by the ongoing genome sequencing efforts in both species. Sixty-three pairs of the sequenced BAC clones analyzed shared an average of nine microsyntenic gene pairs (Choi et al., 2004b). Results from detailed analysis of 10 of the 63 clone pairs with broadly spaced genetic positions in the two genomes showed that approximately 82% of identified genes were syntenic between *M. truncatula* and *L. japonicus*. Tandem duplication accounts for a 12% and a 17% increase in the number of predicted genes in *L. japonicus* and *M. truncatula*, respectively, with only one case that the same homolog duplicated in both species. This observation suggests that the majority of tandem duplication events occurred independently after the divergence of the two species. Intriguingly, in many cases, the regions that were conserved between legumes were also conserved with one or more regions of the Arabidopsis genome, despite at a lower level.

CROSS-SPECIES GENE PREDICTION AND ISOLATION

The conserved genome structure between *M. truncatula* and crop legumes has allowed for map-based cloning of genes required for nodulation in crop legumes, using *M. truncatula* as a surrogate genome (Endre et al., 2002; Limpens et al., 2003). One example is a nodulation receptor kinase (*NORK*) gene that is required for both bacterial and fungal symbiosis (Endre et al., 2002). Three loci with similar nonnodulation mutant phenotypes were mapped to syntenic locations of *M. truncatula*, pea, and alfalfa. The closely linked flanking markers in alfalfa were used as probes to pull out *M. truncatula* BACs that cover the orthologous locus in *M. truncatula*. Map-based cloning and a complementation test were performed in *M. truncatula* and eventually led to the simultaneous cloning of three orthologous genes (i.e. *DOES NOT MAKE INFECTION2* [*DMI2*] in *M. truncatula*, *NORK* in alfalfa, and *SYM19* in pea). At the same time, the ortholog of *L. japonicus* (called symbiosis receptor-like kinase, or *SYMRK*), which is located in a syntenic region of *M. truncatula*, alfalfa, and pea, also was isolated (Stracke et al., 2002, 2004).

A similar strategy also was successful for cloning the pea *SYM2* orthologous genes of *M. truncatula* (Limpens et al., 2003). Pea *SYM2* is a putative Nod-factor entry receptor involved in the rhizobial infection process (Geurts et al., 1997). Map-based cloning of *SYM2* in pea was difficult due to its large genome and the lack of efficient transformation methods. However, the pea *SYM2* region is highly syntenic with *M. truncatula* (Gualtieri et al., 2002). The tightly linked markers flanking the *SYM2* in pea were used to identify *M. truncatula* BACs, and a physical contig (approximately 300 kb) covering the *SYM2* orthologous region of *M. truncatula* was sequenced to identify candidate genes. Using the RNA interference reverse genetic tool, Limpens et al. (2003) showed that two LysM-domain receptor kinases were specifically in-

volved in infection thread formation, and, therefore, are potential orthologs of the *SYM2* in pea.

LEGUME-ARABIDOPSIS COMPARISON: IMPLICATION OF CORRELATED DIVERGENCE OF GENOME STRUCTURE AND FUNCTION?

The degree of conservation of genome structure between legumes and Arabidopsis is less straightforward. Grant et al. (2000) reported substantial macrosynteny between soybean and Arabidopsis, while comparison between *M. truncatula* and Arabidopsis revealed a lack of extended macrosynteny between the two genomes (Zhu et al., 2003). Nevertheless, it is obvious that synteny is frequently maintained over small chromosomal segments. In cases of localized synteny, genetically linked loci in *M. truncatula* often are collinear with several segments of Arabidopsis, consistent with the fact that the Arabidopsis genome has experienced extensive segmental duplication and reshuffling accompanied by selective gene loss (Vision et al., 2000; Bowers et al., 2003). Sequence analyses also revealed networks of microsynteny that are often highly degenerate, similar to that reported by Ku et al. (2000). The erosion of microsynteny could be ascribed to either the selective gene loss from duplicated loci or the absence of close homologs of legume genes in Arabidopsis.

The divergence of genome microstructure has been widely documented, but it is unknown whether such divergence contributes to the divergence of genome function. Comparisons of regions comprising genes responsible for species-specific or family specific phenotypes provide a unique opportunity to answer this question. As described above, the *NORK* orthologs required for nodulation are located in the syntenic regions of the four legume species (i.e. *M. truncatula*, alfalfa, *L. japonicus*, and pea). Comparative sequence analysis of the *M. truncatula* and *L. japonicus* regions revealed highly conserved microsynteny, with 9 of the 11 predicted genes being conserved over an approximately 130-kb interval (Fig. 3). Seven of the 11 distinct genes from the *MtDMI2/LjSYMRK* regions also exhibited microsynteny with four segments of the Arabidopsis genome. The individual Arabidopsis syntenic regions have experienced significant genic rearrangement, with less than four genes in a block being conserved with *M. truncatula* and *L. japonicus*. Nevertheless, a combination of 10 distinct Arabidopsis genes from the four duplicated segments is maintained from a total of 14 distinct genes predicted from the syntenic segments of all three species, identical to the numbers observed in *M. truncatula* and *L. japonicus*. Interestingly, the ortholog of *MtDMI2/LjSYMRK* is missing in the syntenic segments of Arabidopsis. In particular, a legume lectin gene was amplified in syntenic regions of both *M. truncatula* and *L. japonicus*, but not in the Arabidopsis regions. In fact, the syntenic counterparts of the legume lectin genes in Arabidopsis are lectin-like protein kinases comprising both a lectin domain and a Ser/Thr kinase domain, suggesting that domain shuffling might have occurred during the divergence of gene structure among plant genomes. Plant lectins have been implicated as playing an important role in mediating recognition and specificity in the Rhizobium-legume nitrogen-fixing symbiosis (Hirsch, 1999). It is unknown whether such gene amplification in legumes has any particular role in nodulation and nitrogen fixation.

The fact that *NORK* orthologs are extremely conserved in terms of sequence similarity (87%–97%), function, and genomic location among multiple legumes suggests that such divergence of genome microstructure has occurred before the divergence of legume family. The two closest homologs of *MtDMI2/LjSYMRK* in Arabidopsis are At1g67720 and At2g37050 with a sequence identity of approximately 33%. A TBLASTn search of the *M. truncatula* Gene Index using the sequences of At1g67720 and At2g37050 identified highly conserved genes from *M. truncatula*, suggesting that At1g67720 and At2g37050 likely are not the ancient orthologs of *MtDMI2/LjSYMRK*. Therefore, *MtDMI2/LjSYMRK* likely is absent in the lineages giving rise to Arabidopsis. Similar divergence of genome microstructure also was observed in comparisons of maize (*Zea mays*), sorghum (*Sorghum bicolor*), and rice (*Oryza sativa*), where a cluster of maize *zein* storage protein genes are conserved with those of sorghum (*kafirin*), while the homologs of the storage protein genes are completely missing from the rice orthologous segment and elsewhere of the rice genome (Song et al., 2002). These observations indicate that selective gene loss/retention have played an important role in the divergence of species-specific or family specific phenotypes (e.g. nodulation in legumes).

Another legume receptor-like kinase gene required for regulation of the root nodule number recently was cloned from soybean (*GmNARK* for *G. max* nodule autoregulation receptor kinase; Searle et al., 2003) and *L. japonicus* (*LjHAR1* for hypernodulation aberrant roots; Krusell et al., 2002; Nishimura et al., 2002). The putative *M. truncatula* ortholog *MtSUNN* (supernumerary nodules) also was identified and mapped to the syntenic regions of *GmNARK/LjHAR1* (Penmetsa et al., 2003; Schnabel et al., 2003; Choi et al., 2004a). Similar to that observed in the *MtDMI2/LjSYMRK* regions, the *GmNARK/LjHAR1* regions can be aligned with at least seven duplicated segments of the Arabidopsis genome, but none of them contain the *GmNARK/LjHAR1* ortholog. The ortholog also was missing in a paralogous region of *L. japonicus*. In this case, however, the closest homolog of *GmNARK/LjHAR1*, *AtCLAVATA1*, is located in a nonsyntenic region of Arabidopsis. Despite the uncertainty of their orthology, *AtCLAVATA1* and *GmNARK/LjHAR1* share both highly conserved sequence similarity and gene structure, indicating that they are descendants of a common ancestor (Krusell et al., 2002; Nishimura et al., 2002; Searle et al., 2003). Apparently, these

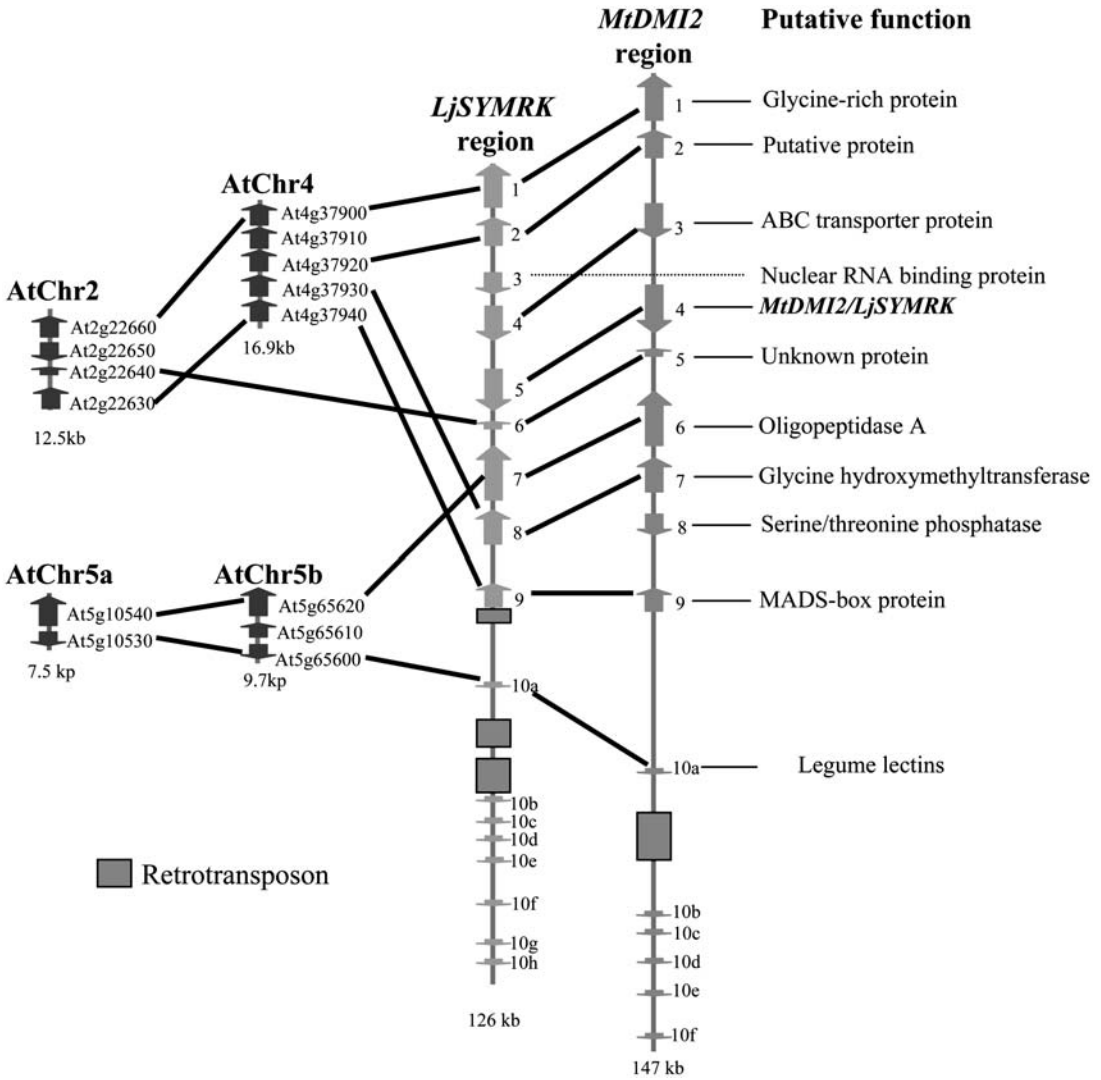

Figure 3. Microsynteny of the *MtDMI2/LjSYMRK* regions and comparison with four segments of Arabidopsis. Lines indicate significant homology matches between predicted genes. The orientations of predicted genes are indicated by arrows. The maps are drawn to scale.

homologous genes have experienced functional divergence between plant families, with *AtCLAVATA1* controlling stem cell proliferation in Arabidopsis, whereas *GmNARK/LjHAR1* functions in the leaf and exerts long-distance control of nodulation (Searle et al., 2003). Such functional divergence appears to have been associated with its relocation in either or both of the lineages leading to Arabidopsis or legumes, and this event should have occurred before the divergence of legume family. As there are two copies of *AtCLAVATA1*-like genes in soybean (*GmCLAVATA1-A* and *GmCLAVATA1-B*; Yamamoto et al., 2000), it was suggested that *GmNARK* (*GmCLAVATA1-B*) is likely a duplicated version of At*CLAVATA1* and has been recruited into nodulation function in legumes (Searle et al., 2003). The recruitment of the *GmNARK* into its current roles is associated with changes in gene expression, leading to differential tissue-specific expression to *AtCLAVATA1* and function in controlling the nodule number (Searle et al., 2003). Alternatively, the ancestor of *CLAVATA1* might be bifunctional, and the descendent duplicated genes evolve to partition the ancestral function by showing organ- and/or time-specific expression (Adams et al., 2003). Thus, losing one copy of the duplicated and subfunctionalized gene during speciation eventually may lead to species-specific phenotypes.

WHAT MAKES A LEGUME?

The divergence of plant phenotypes that distinguish one species from another is a big puzzle for plant biologists. In the case of legumes, an interesting ques-

tion we often ask is as follows: Are genes required for nodulation and nitrogen fixation legume specific? Recent studies have shown that some of the genes required for nodulation (e.g. *DMI1* and *DMI3* in *M. truncatula*) are highly conserved between legumes and non-legumes, and their orthologs can be unambiguously defined in non-legumes such as rice and/or Arabidopsis through microsyntenic analysis (Ane et al., 2004; Levy et al., 2004). These results suggested that at least some of the genes involved in nodulation are evolved from broadly conserved pathways of plant development (Parniske and Downie, 2003; Szczyglowski and Amyot, 2003). This view was further supported by Fedorova et al. (2002), who identified 340 tentative consensus sequences with nodule-specific expression patterns, approximately 40% of which shared sequence homology to sequences from non-legumes. It will be a challenging task to undertake the discovery and dissection of the function of the genes that are required for nodulation but are conserved across plant families (e.g. *DMI1* and *DMI3*) in non-legumes.

In addition to the recruitment of broadly conserved genes for novel legume functions, legumes also may have evolved novel genes that are involved in legume-specific functions. By comparing unigene sets from *M. truncatula*, *L. japonicus*, and soybean to non-legume unigene sets and to the genomic sequences of rice and Arabidopsis, Graham et al. (2004) were able to identify more than 2,500 legume-specific EST contigs, accounting for approximately 6% of genes in legume unigene sets. Motif analysis identified three major gene families from this putative legume-specific gene set, including F-box-related proteins, Pro-rich proteins, and Cys cluster proteins. In particular, the more than 300 Cys cluster proteins, with predicted similarity to defensins, represent approximately 1% of expressed genes in Medicago, primarily from nodules (Fedorova et al., 2002; Mergaert et al., 2003; Graham et al., 2004). Work is under way to characterize these putative legume-specific genes using both forward and reverse genetic tools (K. VandenBosch, personal communication).

CONCLUSION

The past several years have seen tremendous progress in the study of legume genomics, thanks to the development of abundant genetic and genomic resources for two model legumes, *M. truncatula* and *L. japonicus*, and the important crop legume soybean. Undoubtedly, model systems will continue to play a critical role in contributing to our understanding of the mechanisms underlying nodulation and symbiotic nitrogen fixation, the most conserved phenotype in legumes. There is still much to learn about the genomic organization of crop legumes such as soybean, groundnut, and common bean. A major challenge for comparative legume genomics is to translate information gained from model species into improvements in crop legumes. The complexity of that challenge may well be defined by the structural and functional similarities and dissimilarities among these very fascinating genomes.

ACKNOWLEDGMENTS

We thank Dr. Steve Cannon and Dr. Kathryn VandenBosch for their helpful comments on the manuscript. This article (05–06–009) is published with the approval of the Director of the Kentucky Agricultural Experiment Station.

Received December 22, 2004; returned for revision January 18, 2005; accepted January 24, 2005.

LITERATURE CITED

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA **100:** 4649–4654

Ane JM, Kiss GB, Riely BK, Penmetsa RV, Oldroyd GE, Ayax C, Levy J, Debelle F, Baek JM, Kalo P, et al (2004) Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. Science **303:** 1364–1367

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell **12:** 1021–1029

Boutin SR, Young ND, Olson T, Yu ZH, Shoemaker R, Vallejos C (1995) Genome conservation among three legume genera detected with DNA markers. Genome **38:** 928–937

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438

Cannon SB, McCombie WR, Sato S, Tabata S, Denny R, Palmer L, Katari M, Young ND, Stacey G (2003) Evolution and microsynteny of the apyrase gene family in three legume genomes. Mol Genet Genomics **270:** 347–361

Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, et al (2004a) A sequence-based genetic map of Medicago truncatula and comparison of marker colinearity with M. sativa. Genetics **166:** 1463–1502

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, et al (2004b) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA **101:** 15289–15294

Cook DR (1999) *Medicago truncatula*—a model in the making! Curr Opin Plant Biol **2:** 301–304

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell **12:** 637–646

Dixon RA, Sumner LW (2003) Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol **131:** 878–885

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol **131:** 900–910

Endre G, Kereszt A, Kevei Z, Mihacea S, Kalo P, Kiss GB (2002) A receptor kinase gene regulating symbiotic nodule development. Nature **417:** 962–966

Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP (2002) Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula.* Plant Physiol **130:** 519–537

Foster-Hartnett D, Mudge J, Larsen D, Danesh D, Yan H, Denny R, Penuela S, Young ND (2002) Comparative genomic analysis of sequences sampled from a small region on soybean (*Glycine max*) molecular linkage group G. Genome **45:** 634–645

Geurts R, Heidstra R, Hadri AE, Downie JA, Franssen H, Van Kammen A, Bisseling T (1997) Sym2 of pea is involved in a nodulation factor-perception mechanism that controls the infection process in the epidermis. Plant Physiol **115:** 351–359

Graham MA, Silverstein KA, Cannon SB, VandenBosch KA (2004)

Computational identification and characterization of novel genes from legumes. Plant Physiol **135:** 1179–1197

Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. Plant Physiol **131:** 872–877

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots: genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. Proc Natl Acad Sci USA **97:** 4168–4173

Gualtieri G, Kulikova O, Limpens E, Kim DJ, Cook DR, Bisselin T, Geurts R (2002) Microsynteny between pea and Medicago truncatula in the SYM2 region. Plant Mol Biol **50:** 225–235

Hirsch AM (1999) Role of lectins (and rhizobial exopolysaccharides) in legume nodulation. Curr Opin Plant Biol **2:** 320–326

Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis TH, Kiss GB (2004) Comparative mapping between Medicago sativa and Pisum sativum. Mol Genet Genomics **272:** 235–246

Krusell L, Madsen LH, Sato S, Aubert G, Genua A, Szczyglowski K, Duc G, Kaneko T, Tabata S, De Bruijn F, et al (2002) Shoot control of root development and nodulation is mediated by a receptor-like kinase. Nature **420:** 422–426

Ku HM, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and Arabidopsis genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proc Natl Acad Sci USA **97:** 9121–9126

Lee JM, Bush A, Specht JE, Shoemaker RC (1999) Mapping duplicate genes in soybean. Genome **42:** 829–836

Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. Arabidopsis as a bridging species to resolve genome duplication events among legumes. Theor Appl Genet **103:** 765–773

Levy J, Bres C, Geurts R, Chalhoub B, Kulikova O, Duc G, Journet EP, Ane JM, Lauber E, Bisseling T, et al (2004) A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science **303:** 1361–1364

Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R (2003) LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science **302:** 630–633

Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science **264:** 421–424

Menancio-Hautea D, Fatokum CA, Kumar L, Danesh D, Young ND (1993) Comparative genome analysis of mungbean (*Vigna radiata* (L.) Wilczek) and cowpea (*V. unguiculata* (L.) Walpers) using RFLP mapping data. Theor Appl Genet **86:** 797–810

Mergaert P, Nikovics K, Kelemen Z, Maunoury N, Vaubert D, Kondorosi A, Kondorosi E (2003) A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. Plant Physiol **132:** 161–173

Nishimura R, Hayashi M, Wu GJ, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M, et al (2002) HAR1 mediates systemic regulation of symbiotic organ development. Nature **420:** 426–429

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, New York

Pagel J, Walling JG, Young ND, Shoemaker RC, Jackson SA (2004) Segmental duplications within the *Glycine max* genome revealed by fluorescence in situ hybridization of bacterial artificial chromosomes. Genome **47:** 764–768

Parniske M, Downie JA (2003) Plant biology: locks, keys and symbioses. Nature **425:** 569–570

Penmetsa RV, Frugoli JA, Smith L, Long SR, Cook DR (2003) Dual genetic pathways controlling nodule number in *Medicago truncatula*. Plant Physiol **131:** 998–1008

Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. Syst Biol (in press)

Polhill RM, Raven PH, eds (1981) Advances in Legume Systematics, Part 1. Royal Botanic Gardens, Kew, UK

Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. Genome **47:** 868–876

Schnabel E, Kulikova O, Penmetsa RV, Bisseling T, Cook DR, Frugoli J (2003) An integrated physical, genetic and cytogenetic map around the sunn locus of Medicago truncatula. Genome **46:** 665–672

Searle IR, Men AE, Laniya TS, Buzas DM, Iturbe-Ormaetxe I, Carroll BJ, Gresshoff PM (2003) Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. Science **299:** 109–112

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, et al (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics **144:** 329–338

Song R, Llaca V, Messing J (2002) Mosaic organization of orthologous sequences in grass genomes. Genome Res **12:** 1549–1555

Stougaard J (2001) Genetics and genomics of root symbiosis. Curr Opin Plant Biol **4:** 328–335

Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K, et al (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature **417:** 959–962

Stracke S, Sato S, Sandal N, Koyama M, Kaneko T, Tabata S, Parniske M (2004) Exploitation of colinear relationships between the genomes of Lotus japonicus, Pisum sativum and Arabidopsis thaliana, for positional cloning of a legume symbiosis gene. Theor Appl Genet **108:** 442–449

Szczyglowski K, Amyot L (2003) Symbiosis, inventiveness by recruitment? Plant Physiol **131:** 935–940

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in Arabidopsis. Science **290:** 2114–2117

Weeden NF, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered **83:** 123–129

Yamamoto E, Karakaya HC, Knap HT (2000) Molecular characterization of two soybean homologs of *Arabidopsis thaliana* CLAVATA1 from the wild type and fasciation mutant. Biochim Biophys Acta **1491:** 333–340

Yan HH, Mudge J, Kim D-J, Shoemaker RC, Cook DR, Young ND (2003) Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. Theor Appl Genet **106:** 1256–1265

Zhu H, Kim DJ, Baek JM, Choi HK, Ellis LC, Kuester H, McCombie WR, Peng HM, Cook DR (2003) Syntenic relationships between *Medicago truncatula* and Arabidopsis reveal extensive divergence of genome organization. Plant Physiol **131:** 1018–1026